

Antworten auf Nonresponse

Schnell, Rainer

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schnell, R. (2008). Antworten auf Nonresponse. *Sozialwissenschaftlicher Fachinformationsdienst soFid*, Methoden und Instrumente der Sozialwissenschaften 2008/1, 11-23. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205141>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Antworten auf Nonresponse¹

Rainer Schnell

Einleitung

Die Ausschöpfungsquoten akademischer und nicht-akademischer Surveys sind in den letzten Jahren stark gesunken (vgl. Abbildung 1)²

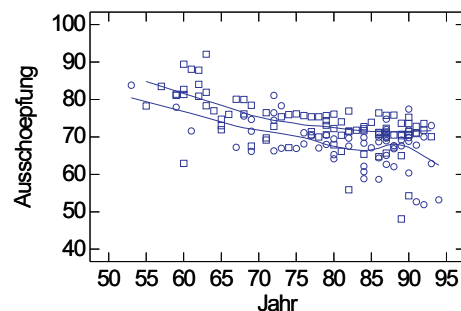


Abbildung 1: Ausschöpfungsraten akademischer und nicht-akademischer Surveys in der BRD

Die - ohnehin nicht unproblematische - Ausschöpfungsquote einer Studie wird häufig als alleiniger Qualitätsindikator für eine Studie gesehen³. Dies ist bekanntlich aus mehreren Gründen irreführend. Abgesehen von den Möglichkeiten der Manipulation der Ausschöpfungsraten – z.B. durch explizit oder implizit zugelassenen Austausch der Zielperson im Haushalt („Proxy-Interviews“) – liegt das Problem der Verwendung von Ausschöpfungsraten als Qualitätsindikator vor allem im Verkennen eines elementaren statistischen Fakts: Für die meisten Statistiken resultiert ein möglicher Nonresponsebias aus dem Produkt des Anteils der Nonrespondenten mit der Differenz zwischen Nonrespondenten und Respondenten. Man benötigt also neben dem Anteil der Nonrespondenten auch Angaben über die Differenz zwischen Nonrespondenten und Respondenten. Erst wenn es hier Unterschiede gibt, ist ein Bias zu erwarten.

-
- 1 Dieser Beitrag ist eine überarbeitete Version eines Vortrags, der auf dem XXXVII. Kongress der deutschen Marktforschung am 7. Mai 2002 in Wolfsburg gehalten wurde.
 - 2 Es handelt sich um eine Analyse der Feldberichte aller Studien, die im Zentralarchiv für empirische Sozialforschung bis 1997 vorhanden waren. Details finden sich bei Schnell (1997). Die untere Kurve (eine Lowess-Glättung) bezieht sich auf akademische Surveys, die obere auf nicht-akademische Surveys (ohne Media-Analysen).
 - 3 Üblicherweise wird die Ausschöpfungsraten als Anteil der realisierten Interviews am bereinigten Stichprobenbrutto definiert (vgl. Schnell 1997:19-27). Problematisch ist hierbei der Begriff der Bereinigung. Bei einigen Stichproben (vor allem bei Varianten des Random-Walks) lässt sich zudem kein Brutto sinnvoll definieren. Aus diesen Gründen sind viele berichtete Ausschöpfungsraten fälschlich überhöht.

Der Bias ist umso höher, je größer der Unterschied zwischen den Respondenten R und den Nichtrespondenten N ist. Weiterhin ist der Bias umso höher, je mehr Nonrespondenten es gibt. Der Nonresponsebias ist also ein Produkt aus der Differenz zwischen Respondenten und Nichtrespondenten und dem Anteil der Nonrespondenten an allen Befragten:

$$Bias_{\bar{x}} = (\bar{x}_R - \bar{x}_N) * \frac{n_N}{n_N + n_R}$$

Die Abbildung 2 zeigt den Bias eines Mittelwerts als Funktion der Nonresponserate zwischen 0 und 100% sowie als Funktion der Differenz zwischen Respondenten und Nichtrespondenten. Wie man sieht, ist ein Bias von Null trotz großer Anteile von Nonresponse auch dann erreichbar, wenn es keine Unterschiede zwischen Respondenten und Nonrespondenten gibt.

Probleme entstehen dann, wenn solche Unterschiede groß und die Nonresponseraten hoch sind.

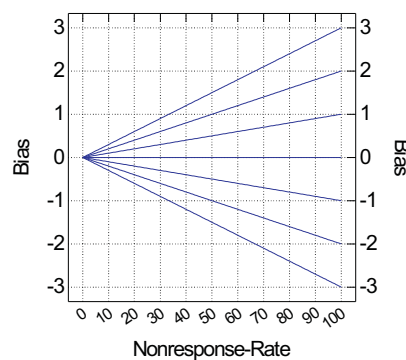


Abbildung 2: Bias als Funktion der Nonresponserate bei metrischen Variablen

Die Abbildung 3 zeigt die am Beispiel von Prozentsatzunterschieden noch einmal. Unterscheiden sich Respondenten und Nichtrespondenten um z.B. 30% und liegt die Ausschöpfung der Stichprobe bei 60%, dann liegt der Bias bei $(1-0.6) * 0.3 = 0.12$.

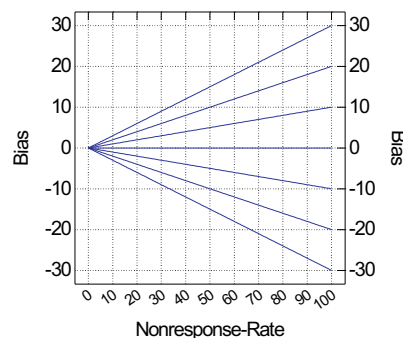


Abbildung 3: Bias als Funktion der Nonresponserate bei Anteilswerten

Die Schätzung auf der Basis der Stichprobe liegt also um 12% zu niedrig. Ob 12% „wenig“ oder „viel“ ist, hängt von der konkreten Fragestellung ab.

Um das potentielle Problem einer geringen Ausschöpfungsrate beurteilen zu können, müssen zunächst verschiedene Formen von Nonresponse unterschieden werden:

Üblicherweise wird zwischen Item-Nonresponse und Unit-Nonresponse unterschieden, wobei Unit-Nonresponse bedeutet, dass alle Angaben für die Zielperson fehlen. Item-Nonresponse bezieht sich auf einzelne Angaben der Zielpersonen. Sowohl bei Item- als auch bei Unitresponse müssen die einzelnen Ursachen für das Fehlen der Daten unterschieden werden. Bei Unit-Nonresponse ist die klassische Unterscheidung zwischen

- Nichterreichbarkeit
- Erkrankung
- Verweigerung

hilfreich. Jeder dieser Formen von Nonresponse liegen andere Mechanismen zugrunde, die sich bei jedem Datenerhebungsmodus und jedem Untersuchungsthema anders auswirken können.

Verweigerungen

Üblicherweise lösen vor allem Verweigerungen Besorgnis aus. Dies steht in bemerkenswerten Gegensatz zur wissenschaftlichen Literatur: Verweigerungen sind in sehr hohem Ausmaß situativ bedingt. Daher gibt es nur schwache Korrelationen zwischen Verweigerungsverhalten und Hintergrundvariablen. Insbesondere gibt es keine empirischen Hinweise auf einen „harten Kern“, der sich jedem Befragungsversuch entzieht. Konvertierungen von Verweigerern sind anscheinend prinzipiell möglich. Ob sich der Aufwand hierfür lohnt ist hingegen m.E. empirisch nicht geklärt.

Erkrankung

Zumeist ein vergleichsweise kleiner Anteil der Befragten fällt durch Erkrankung aus. Hierbei handelt es sich zum kleineren Teil um verdeckte Interviewer-, „Verweigerungen“ oder stellvertretende Verweigerungen. Ob sich der verbleibende Rest auf die Ergebnisse auswirkt, hängt vom Thema der Untersuchung ab. Sobald das Thema mit Variablen zum Gesundheitsverhalten korreliert, besteht die Gefahr erheblicher Fehlschlüsse selbst bei kleinen Anteilen von Erkrankten.

Als Beispiel kann die Berechnung eines Krankheitsindex anhand der Daten des SOEP dienen⁴. Wie man der Abbildung entnehmen kann, gibt es deutliche Unterschiede im Gesundheitszustand während der Welle a in Abhängigkeit vom Bearbeitungsergebnis der Welle b. Die Teilnehmer der Welle b sind daher im Vergleich zu den Teilnehmern der Welle a tendenziell gesünder. Anders gesagt: Eine Querschnittsanalyse auf der Basis der Welle B würde den Gesundheitszustand der Bevölkerung überschätzen.

4 Vgl. Heller/Schnell (2000). Der KEI für die Person i ist wie folgt definiert: $KEI_i = \text{Anzahl Arztbesuche}_i * \text{Gewicht} + \text{Anzahl Krankenhausnächte}_i$. (Fuchs und Hansmeier 1996). Das Gewicht dient zur Korrektur bei chronisch Kranken.

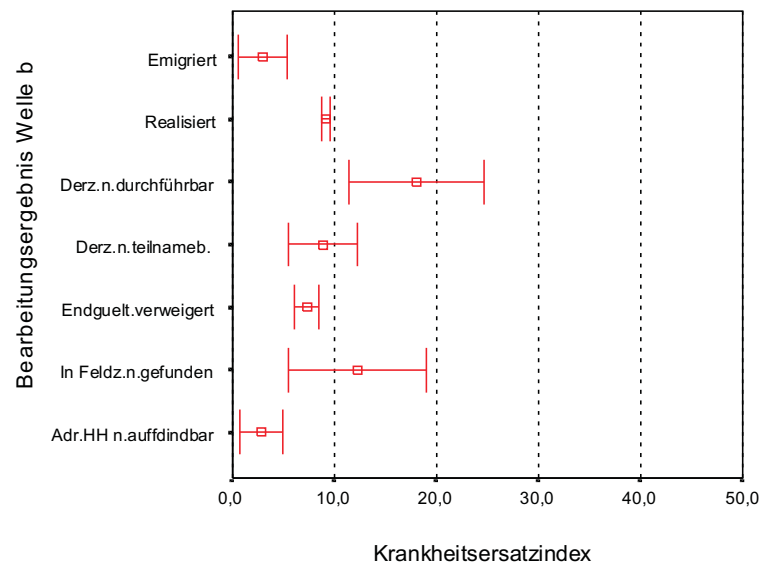
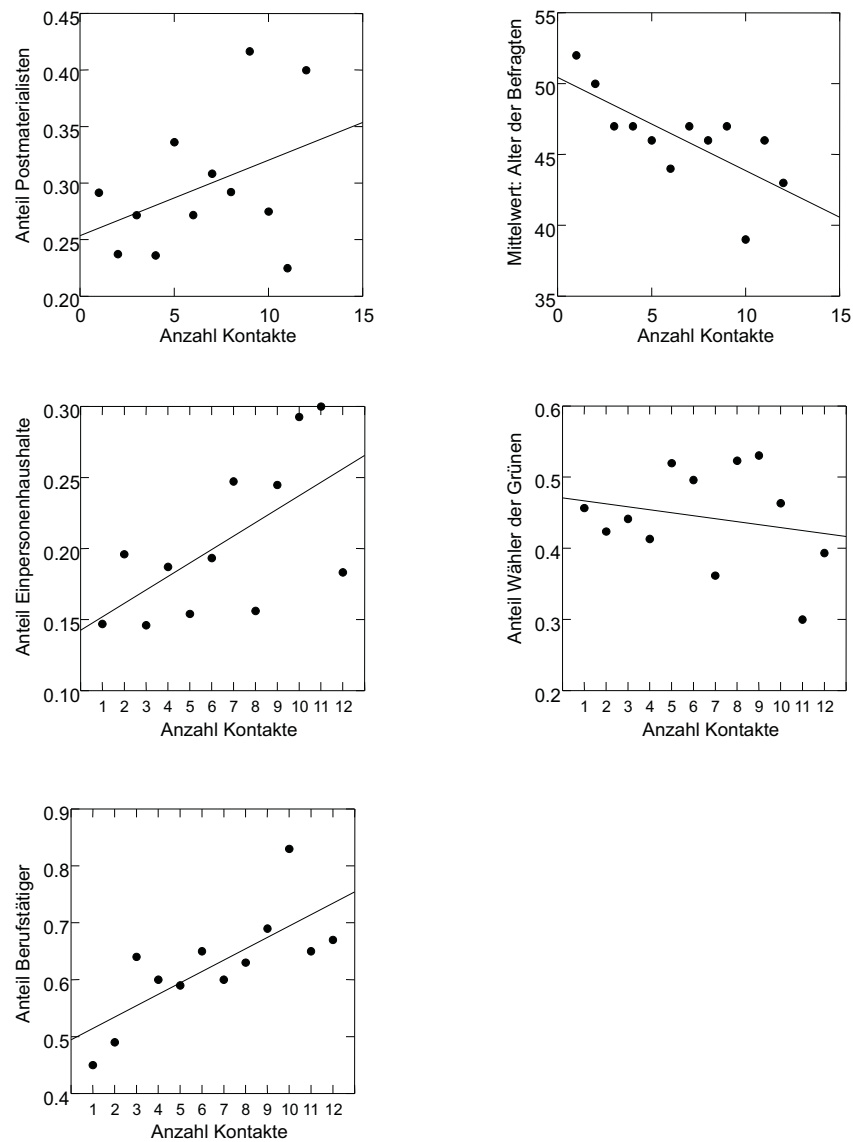


Abbildung 4: 95%-Konfidenzintervalle für den Krankheitsersatzindex in Welle a nach Bearbeitungsergebnis in Welle b (SOEP, Westdeutsche, ohne Verstorbene 1984-85)

Nichterreichte

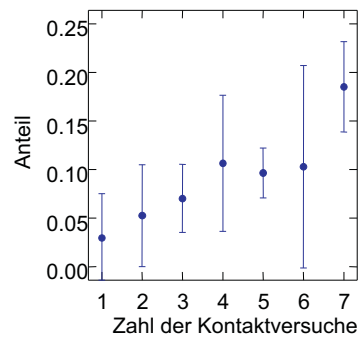
Nichterreichbarkeit korreliert mit sehr vielen markt- und sozialwissenschaftlich relevanten Variablen, so mit Erwerbstätigkeit, Zahl der Personen im Haushalt, Alter, Postmaterialismus dem Anteil der Grün-Wähler usw. (Abbildungen 4a-4e).

Schwererreichbare können nicht einfach durch leichter Erreichbare ersetzt werden: Auch Schwererreichbare sind keine homogene Gruppe. Die Effekte der Erreichbarkeit variieren je nach den aktuellen Feldbedingungen selbst innerhalb des gleichen Surveys. Die lässt sich anhand einer Reihe von Variablen anhand der Daten des Defect-Projekts (Schnell/Kreuter 2000, Abbildungen 5a-) leicht zeigen.

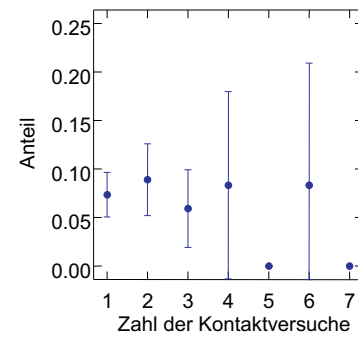


Abbildungen 4a-e: Erreichbarkeit und sozialwissenschaftlich relevante Variablen im ALLBUS 2000 West

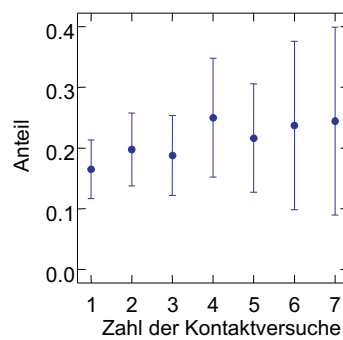
Wohnungseinbruch – Institut 1



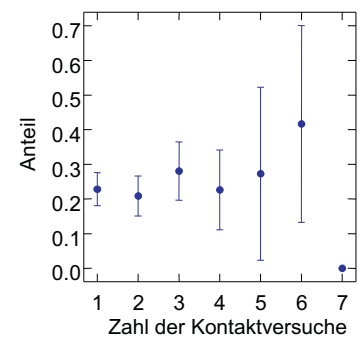
Wohnungseinbruch – Institut 2



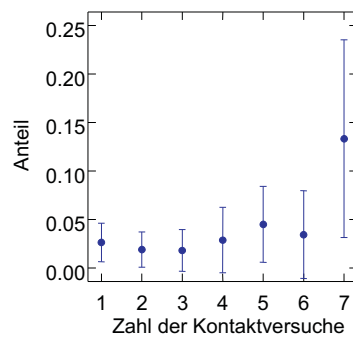
Autoaufbruch – Institut 1



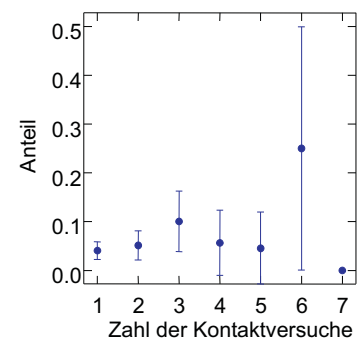
Autoaufbruch – Institut 2



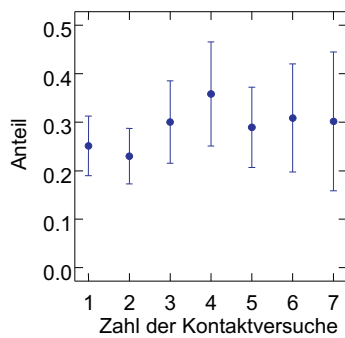
Autodiebstahl – Institut 1



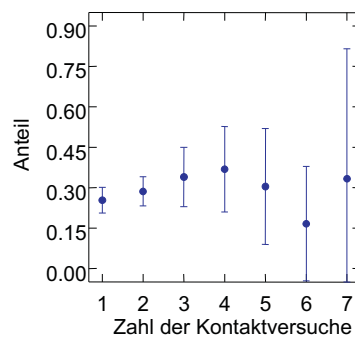
Autodiebstahl – Institut 2

**Abbildungen 5a-f:** Erreichbarkeit und sozialwissenschaftlich relevante Variablen in den DEFECT-Surveys

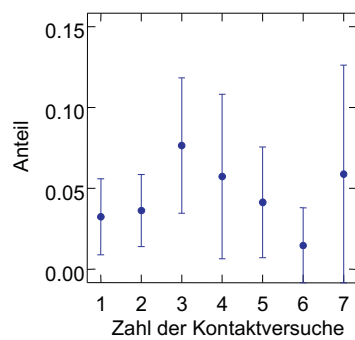
Fahrraddiebstahl – Institut 1



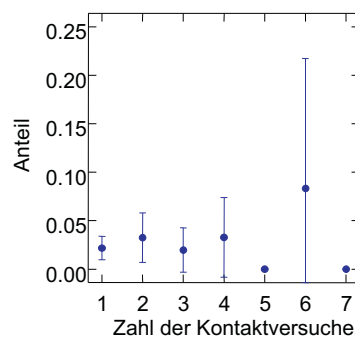
Fahrraddiebstahl – Institut 2



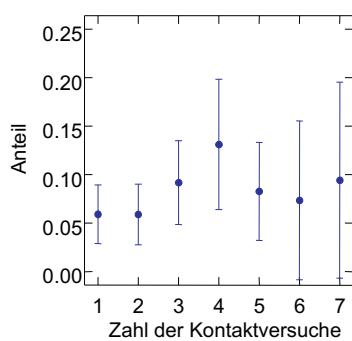
Opfer eines Raubüberfalls – Institut 1



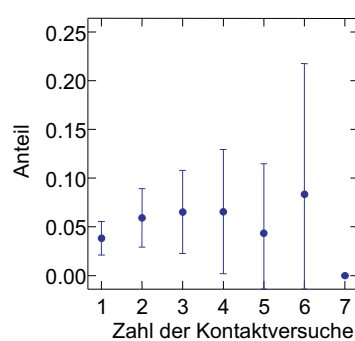
Opfer eines Raubüberfalls – Institut 2



Von fremder Person geschlagen – Institut 1

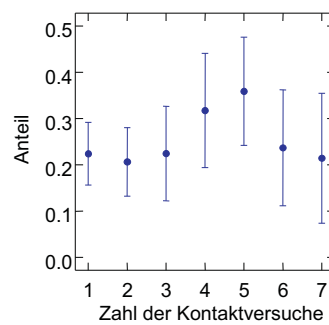


Von fremder Person geschlagen – Institut 2

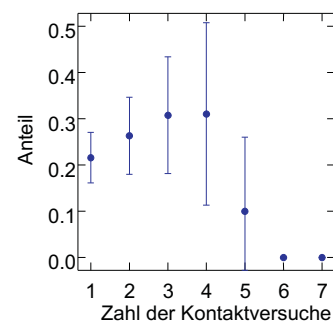


Abbildungen 5g-l: Erreichbarkeit und sozialwissenschaftlich relevante Variablen in den DEFECT-Surveys

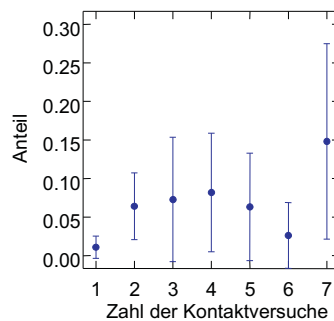
Sexuelle Belästigung am Telefon – Institut 1



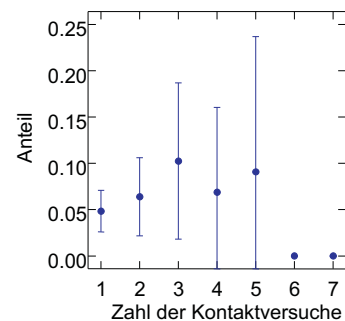
Sexuelle Belästigung am Telefon – Institut 2



Opfer eines sexuellen Angriffs – Institut 1



Opfer eines sexuellen Angriffs – Institut 2

**Abbildungen 5m-p:** Erreichbarkeit und sozialwissenschaftlich relevante Variablen in den DEFECT-Surveys

Die Ergebnisse dieser Studie legen es ebenfalls nahe, dass sich die Ergebnisse der Feldarbeit nach der Art der Kontaktaufnahme (telefonisch oder persönlich) selbst innerhalb einer Face-to-Face-Befragung unterscheiden. Entsprechend benötigt man für eine eventuelle Korrektur auf der Basis der Erreichbarkeit die tatsächlichen Kontaktzahlen in allen Kontaktmodi (telefonisch, schriftlich, persönlich). Als Ersatz wird für diese Daten häufig auf die Angaben der Befragten über ihre Erreichbarkeit zurückgegriffen. Die Erinnerung bzw. der Bericht der Befragten über ihre Erreichbarkeit ist zu fehlerhaft um als Basis der Korrektur zu dienen. Das gilt auch für die Erinnerung bzw. die Berichte der Interviewer. So lagen die Korrelationen zwischen berichteter Zahl der Abende, die die Befragten zu Hause verbrachten, und der Zahl der tatsächlichen Kontaktversuche bei zwei identischen Surveys bei einem Institut bei 0.07, bei dem anderen Institut bei -0.12^5 .

5 Quelle: eigene Berechnungen auf der Basis der DEFECT-Surveys I+II, vgl. Schnell/Kreuter (2000).

Korrekturverfahren

Sinnvollerweise unterscheidet man verschiedene Typen von Korrekturverfahren. Zunächst kann man bei den älteren Verfahren unterscheiden zwischen

- Oversampling
- Gewichtung durch Angleichen der Randverteilung

Bei den neueren Verfahren ist die Unterscheidung zwischen

- Sample-Selection-Modellen
- Propensity Weighting
- Imputationen
- Multiple Imputationen

sinnvoll.

Ältere Korrekturverfahren

In der Praxis der Markt- und Sozialforschung werden üblicherweise nur die älteren Verfahren Oversampling und Randverteilungsgewichte verwendet. Beide Verfahren leiden unter dem gleichen Problem. Es müssen – nachweisbar falsche - Homogenitätsannahmen gemacht werden: Innerhalb einer Gewichtungszelle müssen die Ausfälle unsystematisch sein. Entsprechend kann man zeigen, dass sich systematische Ausfallmechanismen durch Gewichtungsverfahren nur sehr bedingt korrigieren lassen.

Die Abbildung 6 zeigt die Veränderung des Mittelwerts des Postmaterialismusindex durch Löschen der Interviews, die jeweils erst nach 2, 3, 4 usw. Kontaktversuchen realisiert werden konnten⁶. Jeder der reduzierten Datensätze wurde jeweils mit den Variablen Alter, Geschlecht und Bundesland neu gewichtet (Schnell 1993:25). Jeder gewichtete Datensatz entspricht also exakt der Sollvergabe der Kombination Alter, Geschlecht und Bundesland. Anders gesagt: Jeder Datensatz entspricht exakt den Bevölkerungsanteilen in den entsprechenden Zellen. Trotzdem zeigen sich signifikante Abweichungen von den tatsächlichen Werten. Solche Veränderungen trotz Gewichtung zeigt sich auch in Regressionsmodellen: Ein Regressionsmodell zur Vorhersage des Postmaterialismusindex durch Alter und Bildung zeigt u.a. eine Veränderung der Regressionskoeffizienten für Alter durch den Ausfallmechanismus trotz Gewichtung jedes reduzierten Datensatzes. Trotz ihrer theoretischen und praktischen Nachteile sind Gewichtungsverfahren im Gegensatz zu allen anderen Verfahren aber in der Praxis recht populär, da keine Brutto-Datensätze benötigt werden.

6 Datenbasis: Allbus 1980. Die Abbildungen wurden gegenüber dem Original geringfügig modifiziert (Skalierungen, Label etc.).

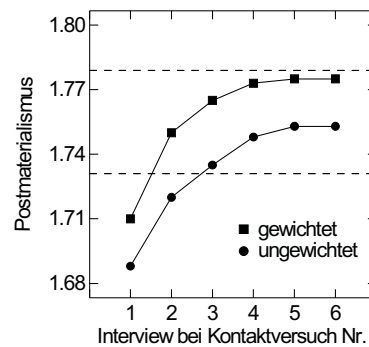


Abbildung 6: Veränderung im Postmaterialismus-Index trotz Gewichtung bei Beschränkung auf Leichterreichbare

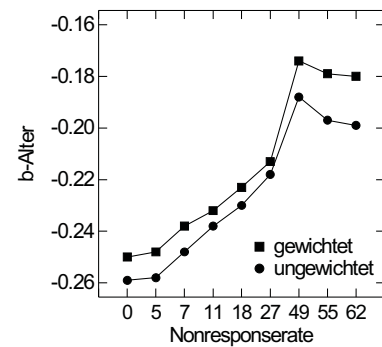


Abbildung 7: Veränderung des Regressionskoeffizienten für Alter im Regressionsmodell zur Erklärung

Neuere Korrekturverfahren

Sample-Selection-Modelle („Heckman-Korrekturen“) basieren auf der Spezifikation einer Modellgleichung und einer Selektionsgleichung. Über Annahmen über die Korrelation der Fehlerterme kann – falls die Gleichungen empirisch erfüllt sind – die Modellgleichung korrekt geschätzt werden⁷. Diese Korrekturen eignen sich nur für die Schätzung von Regressionsparametern unter der Annahme korrekt spezifizierter Gleichungen. Anders gesagt: Man benötigt eine korrekte inhaltliche Theorie über den Gegenstandsbereich. Die statistischen Annahmen dieser Modelle sind mit den gegebenen Daten prinzipiell unüberprüfbar. In der Praxis außerhalb der Ökonometrie spielen diese Korrekturen daher keine Rolle. Simulationsstudien lassen die Ergebnisse dieser Modelle eher skeptisch erscheinen (vgl. Stolzenberg/Relles 1997).

„Propensity Weighting“ basiert typischerweise auf der Schätzung der Teilnahmebereitschaft bzw. der Erreichbarkeit durch eine logistische Regression. Die vorhergesagten „Wahrscheinlichkeiten“ werden dann für eine Gewichtung verwendet (Rosenbaum/Rubin 1983, 1984). Trivialerweise setzt eine Korrektur durch ein solches Verfahren eine korrekt spezifizierte Selektionsgleichung voraus. Generell kann festgehalten werden, dass die Auswahl der Variablen für ein solches Modell wichtiger ist als die exakte Form des Modells (Rizzo/Kalton/Brick 1994). Weiterhin benötigt man Angaben über die Nonrespondenten. Solche Modelle werden in der Praxis überwiegend nur für Panelstudien verwendet, z.B. innerhalb des SOEP. Zwar liegen die diagnostischen Methoden für solche Modelle auf der Hand (Eltinge 2002) werden aber in der Praxis kaum angewandt. Beurteilt man den tatsächlichen Modellfit solcher Verfahren, so ist das Ergebnis enttäuschend (Heller/Schnell 2000).

„Imputationen“ bestehen aus der Ersetzung fehlender Werte durch geschätzte Werte. Typischerweise werden Imputationen eher für Item-Nonresponse verwendet, aber Anwendungen auf Unit-Nonresponse sind möglich. Es gibt zahlreiche Verfahren (Hot-Deck, Mean-Substitution, Regressions-schätzungen), die sich in jeweiligen Annahmen und Ergebnissen unterscheiden (vgl. Schnell 1986).

⁷ Es gibt einige Modellerweiterungen, die andere Annahmen erfordern, vgl. Vella (1998).

Alle diese Verfahren basieren auf der Annahme, dass die Ausfälle zufällig erfolgen (genauer: MCAR, vgl. Little/Rubin 2002:12). Alle einfachen Imputationsverfahren führen zur Unterschätzung der Varianz der geschätzten Statistiken, d.h. imputierte Datensätze haben selbst im günstigsten Fall größere Konfidenzintervalle als es scheint. Multiple Imputationen führen Imputationen mehrfach durch und erlauben so eine korrekte Schätzung der Varianz der geschätzten Statistiken. MI ist eine Monte-Carlo-Simulationsmethode, bei der ein Datensatz nicht einmal ergänzt wird, sondern mehrfach (typischerweise 5 mal). Jeder einzelne Datensatz wird mit Standardmethoden und Standardprogrammen analysiert, die Ergebnisse der multiplen Datensätze werden anschließend kombiniert. Für die Durchführung der multiplen Imputationen existiert eine Reihe von Softwareprogrammen, die z.T. kostenfrei im Netz erhältlich sind⁸. Die Anwendung dieser Techniken erfordert nicht mehr die Annahme rein zufälligen Fehlens (MCAR) sondern nur noch die Annahme, dass die Tatsache, dass Daten fehlen, nicht von unbeobachteten Werten der beobachteten Variablen abhängen. Für die meisten sozialwissenschaftlichen Fragestellungen und Surveys erscheint MAR plausibel. Ausnahmen sind aber möglich, z.B. in medizinischen Surveys oder bei der Analyse seltener Subpopulationen. Leider ist diese Annahme nicht prüfbar; allerdings legen Simulationen die relative Robustheit von MI nahe. Damit besteht der wesentliche „Nachteil“ der MI darin, dass nicht mehr eine Statistik für eine Frage berechnet werden muss, sondern ein Vektor. Man bekommt also nicht mehr „eine“ Antwort, sondern mehrere. Innerhalb der akademischen Statistik dürfte MI daher zukünftig Standard werden⁹. Praktisch wird die Einführung auf Probleme stoßen. Entsprechend gibt es bislang weltweit nur einen Datensatz, der als MI-imputierter Datensatz geliefert wird: NHANES III¹⁰.

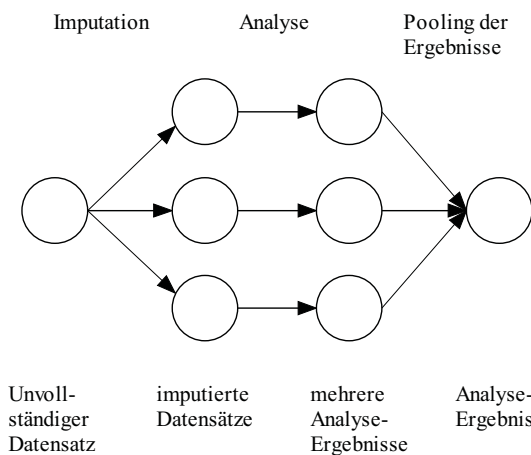


Abbildung 8: Multiple Imputation¹¹

8 Links auf die Programme finden sich unter <http://www.uni-konstanz.de/FuF/Verwiss/Schnell/mdprogs.html>

9 Vgl. einführend Schafer/Olsen (1998), Schafer, J.L. (1999), Collins/Schafer/Kam (2001) und Schafer/Graham (2002).

10 Einzelheiten finden sich bei Barnard/Meng (1999), vgl. auch <http://www.cdc.gov/nchs/datawh/ftp/ftpdata/ftpdata.htm>

11 Diese Abbildung basiert auf einer Idee von Stef van Buuren (<http://web.inter.nl.net/users/S.van.Buuren/mi/html/whatis.htm>).

Schlussfolgerungen

Nonresponse kann weder vermieden noch ignoriert werden. Die häufig vorgeschlagene Verwendung anderer Stichprobenverfahren (z.B. Quota), Erhebungsverfahren (z.B. Internetsurveys) oder Gewichtungen bieten keine Lösung des Nonresponseproblems, sondern verschleiern das Problem lediglich. Die einzig mögliche Antwort auf Nonresponse kann nur in sorgfältiger Feldarbeit, deren Dokumentation und einer statistischen Analyse, die auf Nonresponse Rücksicht nimmt, bestehen. Neue statistische Entwicklungen erlauben die Abschätzung der Unsicherheit der Schlussfolgerungen durch Nonresponse. Um solche Verfahren anwenden zu können, muss ein Datensatz eine Reihe von Informationen enthalten:

- Interviewer-ID
- Sampling-Point-ID
- Datum und Uhrzeit der Kontaktversuche
- Anzahl aller Kontaktversuche (persönlich, telefonisch, schriftlich)

Diese Informationen sind nahezu kostenneutral zu gewinnen. Dies gilt auch für die wichtigste praktische Schlussfolgerung: Für eine Analyse, die Rücksicht auf Nonresponse nimmt, benötigt man immer Brutto-Datensätze, d.h. Records auch für Nichtbefragte. Ohne diese Brutto-Datensätze sind kaum Nonresponse-Analysen und nur sehr begrenzt eventuelle Korrekturen möglich.

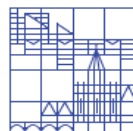
Literatur

- Allison, Paul D. (2002): Missing data, Thousand Oaks (Sage).
- Barnard, J./Meng, X.L. (1999): Applications of multiple imputation in medical studies: from AIDS to NHANES; in: Statistical Methods in Medical Research, 8, 17-36.
- Breen, R. (1996): Regression Models. Censored, Sample Selected or Truncated Data, Thousand Oaks (Sage).
- Collins, L.M./Schafer, J.L./Kam, C.M. (2001): A comparison of inclusive and restrictive missing-data strategies in modern missing-data procedures; in: Psychological Methods, 6, S.330–351.
- Eltinge, J.L. (2002): Diagnostics for the Practical Effects of Nonresponse Adjustment Methods; in: Groves, R.M. u.a. (eds.): Survey Nonresponse, New York (Wiley), S.431-443).
- Fuchs J./Hansmeier, T. (1996). Ein Krankheitsersatzindex: Konstruktion und Validierung; in: Sozial und Präventivmedizin, 41, S.231-239.
- Groves, R. M./Dillman, D. A./Eltinge, J. L./Little, R. J. A. (Eds.) (2002): Survey Nonresponse. New York (Wiley).
- Groves, R. M./Couper, M. P. (1998) : Nonresponse in Household Interview Surveys. New York (Wiley).
- Heller, G./Schnell, R. (2000): The Choir Invisible. Zur Analyse der gesundheitsbezogenen Panelmortalität im SOEP; in: Helmert, U./Bamman, K./Voges, W./Müller, R. (Hrsg.): Müssen Arme früher sterben? Soziale Ungleichheit und Gesundheit in Deutschland, München (Juventa), S.115-134.
- Little, Roderick J.A./Rubin, Donald B. (2002): Statistical Analysis With Missing Data, New York (Wiley), 2. Auflage.

- Rizzo, L./Kalton, G./Brick, J.M. (1994): Nonresponse in the SIPP. Final report, Rockville 1994
- Rosenbaum, P. and Rubin, D.B. (1983): The central role of the propensity score in observational studies for causal effects; in: *Biometrika*, 70, S. 41-55.
- Rosenbaum, P.R. and Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score; in: *Journal of the American Statistical Association*, 79, S. 516-524.
- Schafer, J.L./Olsen, M.K. (1998): Multiple imputation for multivariate missing-data problems: a data analyst's perspective; in: *Multivariate Behavioral Research*, 33, S.545-571.
- Schafer, J.L. (1999): Multiple imputation: a primer; in: *Statistical Methods in Medical Research*, 8, S.3-15.
- Schafer, J.L./Graham, J.W. (2002): Missing data: our view of the state of the art; in: *Psychological Methods*, 7, S.147-177.
- Schnell, R. (1986): Missing-Data-Probleme in der empirischen Sozialforschung, Dissertation, Bochum (Volltext unter <http://www.ub.uni-konstanz.de/kops/volltexte/2000/549>).
- Schnell, R. (1993): Die Homogenität sozialer Kategorien als Voraussetzung für „Repräsentativität“ und Gewichtungungsverfahren; in: *Zeitschrift für Soziologie*, 22, 1, S.16-32.
- Schnell, R. (1997): Nonresponse in Bevölkerungsumfragen, Opladen (Leske + Budrich).
- Schnell, R./Kreuter, F. (2000): Das DEFECT-Projekt: Sampling-Errors und Nonsampling-Errors in komplexen Bevölkerungstichproben; in: *ZUMA-Nachrichten*, 47, 89-101.
- Stolzenberg, R.M./Relles, D.A. (1997): Tools for intuition about sample selection bias and its correction; in: *American Sociological Review*, 62, S.494-507.
- Vella, F. (1998): Estimating Models with Sample Selection Bias: A Survey; in: *Journal of Human Resources*, 33, S.127-169.

Zur Person

Prof. Dr. Rainer Schnell
Universität Konstanz
Professur für Methoden
der empirischen Politik- und Verwaltungsforschung



Kontakt:
Rainer.Schnell@Uni-Konstanz.de